

Genome sequence of the clover-nodulating *Rhizobium leguminosarum* bv. *trifolii* strain SRDI943

Wayne Reeve^{*1}, Elizabeth Drew², Ross Ballard², Vanessa Melino¹, Rui Tian¹, Sofie De Meyer¹, Lambert Brau³, Mohamed Ninawi¹, Hajnalka Daligault^{4,5}, Karen Davenport⁴, Tracy Erkkila⁴, Lynne Goodwin⁵, Wei Gu⁴, Christine Munk⁴, Hazuki Teshima⁴, Yan Xu⁴, Patrick Chain⁴ and Nikos Kyripides⁵

¹Centre for Rhizobium Studies, Murdoch University, Western Australia, Australia

²South Australian Research and Development Institute, Urrbrae, South Australia, Australia

³School of Life and Environmental Sciences, Faculty of Science & Technology, Deakin University, Melbourne, Victoria, Australia

⁴Los Alamos National Laboratory, Bioscience Division, Los Alamos, New Mexico, USA

⁵DOE Joint Genome Institute, Walnut Creek, California, USA

***Correspondence:** Wayne Reeve (W.Reeve@murdoch.edu.au)

Keywords: root-nodule bacteria, nitrogen fixation, rhizobia, *Alphaproteobacteria*

Rhizobium leguminosarum bv. *trifolii* SRDI943 (strain syn. V2-2) is an aerobic, motile, Gram-negative, non-spore-forming rod that was isolated from a root nodule of *Trifolium michelianum* Savi cv. Paradana that had been grown in soil collected from a mixed pasture in Victoria, Australia. This isolate was found to have a broad clover host range but was sub-optimal for nitrogen fixation with *T. subterraneum* (fixing 20-54% of reference inoculant strain WSM1325) and was found to be totally ineffective with the clover species *T. polymorphum* and *T. pratense*. Here we describe the features of *R. leguminosarum* bv. *trifolii* strain SRDI943, together with genome sequence information and annotation. The 7,412,387 bp high-quality-draft genome is arranged into 5 scaffolds of 5 contigs, contains 7,317 protein-coding genes and 89 RNA-only encoding genes, and is one of 100 rhizobial genomes sequenced as part of the DOE Joint Genome Institute 2010 Genomic Encyclopedia for Bacteria and Archaea-Root Nodule Bacteria (GEBA-RNB) project.

Introduction

The availability of usable nitrogen (N) is vital for productivity in agricultural systems that are N-deficient [1]. It can be supplied exogenously in the form of industrially synthesized fertilizers. However, this practice is expensive since fertilizer manufacture depends on the availability of fossil fuels that are burnt to support the industrial process of chemical N-fixation. A far more economical practice is to supply plant-available N to farming systems by exploiting the process of biological N-fixation that occurs in a symbiotic relationship between legumes and their rhizobial microsymbionts [2]. In this specific association, atmospheric inert dinitrogen gas is converted into bioavailable N to support legume growth.

Pasture legumes, including the clovers that comprise the *Trifolium* genus, are major contributors of biologically fixed nitrogen (N₂) to mixed farm-

ing systems throughout the world [3,4]. In Australia, soils with a history of growing *Trifolium* spp. have developed large and symbiotically diverse populations of *Rhizobium leguminosarum* bv. *trifolii* (*R. l. trifolii*) that are able to infect and nodulate a range of clover species. The N₂-fixation capacity of the symbioses established by different combinations of clover hosts (*Trifolium* spp.) and strains of *R. l. trifolii* can vary from 10 to 130% when compared to an effective host-strain combination [5-8].

R. l. trifolii strain SRDI943 (syn. V2-2 [9]) was isolated from a nodule recovered from the roots of the annual clover *Trifolium michelianum* Savi cv. Paradana that had been inoculated with soil collected from under a mixed pasture at Walpeup, Victoria, Australia and grown in N deficient media for four weeks after inoculation, in the greenhouse

[10]. SRDI943 forms an effective symbiosis with *T. purpureum* but sub-optimal N₂-fixation symbiosis with *T. subterraneum* cv. Campeda and Clare (~24 and 54% respectively of that with strain WSM1325 [9,11]). Here we present a preliminary description of the general features for *R. l. trifolii* strain SRDI943 together with its genome sequence and annotation.

Classification and general features

R. l. trifolii strain SRDI943 is a motile, Gram-negative rod (Figure 1 Left and Center) in the order *Rhizobiales* of the class *Alphaproteobacteria*. It is fast growing, forming colonies within 3-4 days when grown on half strength Lupin Agar (½LA) [12] at 28°C. Colonies on ½LA are white-opaque, slightly domed and moderately mucoid with smooth margins (Figure 1 Right).

Minimum information about the Genome Sequence (MIGS) is provided in Table 1. Figure 2 shows the phylogenetic relationship of *R. l. trifolii* strain SRDI943 to root nodule bacteria in the order *Rhizobiales* in a 16S rRNA sequence based tree. This strain clusters closest to *R. l. trifolii* T24 and *Rhizobium leguminosarum* bv. *phaseoli* RRE6 with 100% and 99.8% sequence identity, respectively.

Symbiotaxonomy

R. l. trifolii SRDI943 forms nodules on (Nod⁺) and fixes N₂ (Fix⁺) with a range of annual and perenni-

al clover species of Mediterranean origin (Table 2). SRDI943 forms white, ineffective (Fix⁻) nodules with the perennial clover *T. pratense* and *T. polymorphum*.

Genome sequencing and annotation information

Genome project history

This organism was selected for sequencing on the basis of its environmental and agricultural relevance to issues in global carbon cycling, alternative energy production, and biogeochemical importance, and is part of the Community Sequencing Program at the U.S. Department of Energy, Joint Genome Institute (JGI) for projects of relevance to agency missions. The genome sequence is deposited in the Genomes OnLine Database (GOLD) [33] and an improved-high-quality-draft genome sequence in IMG/GEBA. Sequencing, finishing and annotation were performed by the JGI. A summary of the project information is shown in Table 3.

Growth conditions and DNA isolation

R. l. trifolii strain SRDI943 was cultured to mid logarithmic phase in 60 ml of TY rich media [34] on a gyratory shaker at 28°C. DNA was isolated from the cells using a CTAB (Cetyl trimethyl ammonium bromide) bacterial genomic DNA isolation method [35].

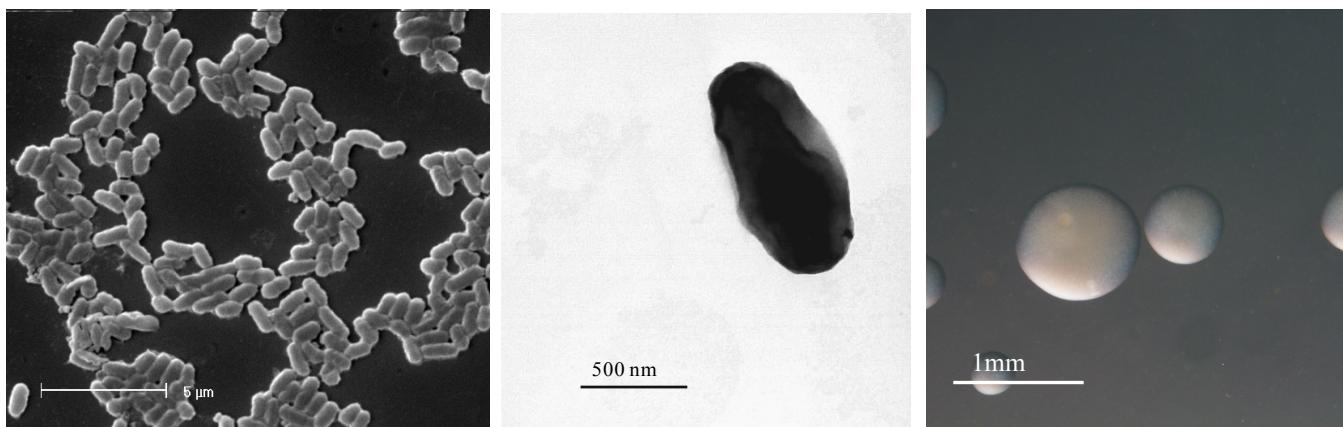


Figure 1. Images of *Rhizobium leguminosarum* bv. *trifolii* strain SRDI943 using scanning (Left) and transmission (Center) electron microscopy as well as light microscopy to show the colony morphology on solid media (Right).

Table 1. Classification and general features of *Rhizobium leguminosarum* bv. *trifolii* SRDI943 according to the MIGS recommendations [13]

MIGS ID	Property	Term	Evidence code
		Domain <i>Bacteria</i>	TAS [14]
		Phylum <i>Proteobacteria</i>	TAS [15]
		Class <i>Alphaproteobacteria</i>	TAS [16,17]
	Current classification	Order <i>Rhizobiales</i>	TAS [17,18]
		Family <i>Rhizobiaceae</i>	TAS [19-21]
		Genus <i>Rhizobium</i>	TAS [21-26]
		Species <i>Rhizobium leguminosarum</i> bv. <i>trifolii</i>	TAS [21,23,27,28]
	Gram stain	Negative	IDA
	Cell shape	Rod	IDA
	Motility	Motile	IDA
	Sporulation	Non-sporulating	NAS
	Temperature range	Mesophile	NAS
	Optimum temperature	28°C	NAS
	Salinity	Non-halophile	NAS
MIGS-22	Oxygen requirement	Aerobic	TAS [11]
	Carbon source	Varied	NAS
	Energy source	Chemoorganotroph	NAS
MIGS-6	Habitat	Soil, root nodule, on host	TAS [9]
MIGS-15	Biotic relationship	Free living, symbiotic	TAS [9]
MIGS-14	Pathogenicity	Non-pathogenic	NAS
	Biosafety level	1	TAS [29]
	Isolation	Root nodule	TAS [9]
MIGS-4	Geographic location	Victoria, Australia	TAS [9]
MIGS-5	Soil collection date	Dec, 1998	IDA
MIGS-4.1	Longitude	142.0262	
MIGS-4.2	Latitude	-35.13531	IDA
MIGS-4.3	Depth	0-10cm	
MIGS-4.4	Altitude	Not recorded	

Evidence codes – IDA: Inferred from Direct Assay; TAS: Traceable Author Statement (i.e., a direct report exists in the literature); NAS: Non-traceable Author Statement (i.e., not directly observed for the living, isolated sample, but based on a generally accepted property for the species, or anecdotal evidence). These evidence codes are from the Gene Ontology project [30].

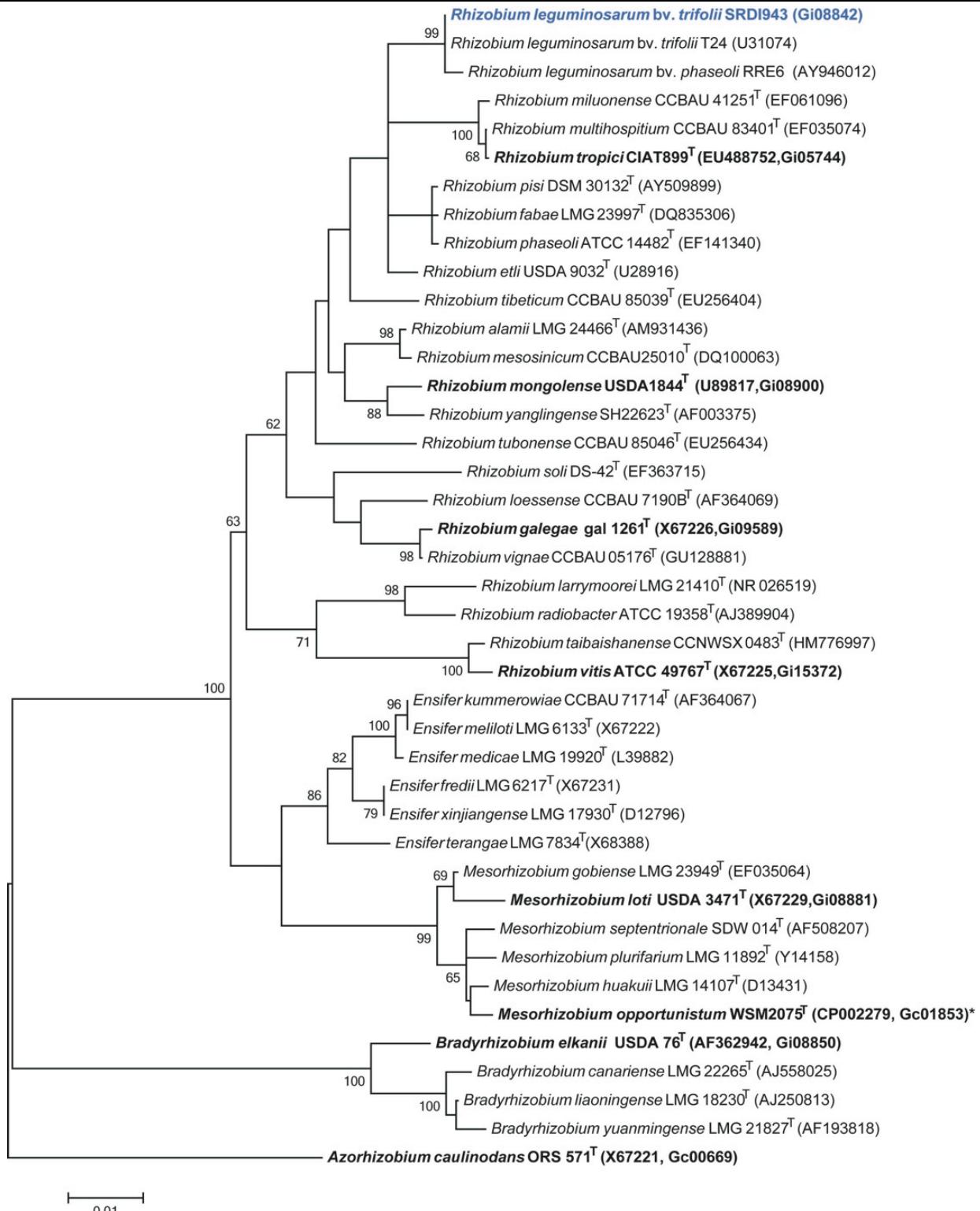


Figure 2. Phylogenetic tree showing the relationship of *Rhizobium leguminosarum* bv. *trifolii* SRDI943 (shown in blue print) with some of the root nodule bacteria in the order Rhizobiales based on aligned sequences of the 16S rRNA gene (1,307 bp internal region). All sites were informative and there were no gap-containing sites. Phylogenetic analyses were performed using MEGA, version 5.05 [31]. The tree was built using the maximum likelihood method with the General Time Reversible model. Bootstrap analysis [32] with 500 replicates was performed to assess the support of the clusters. Type strains are indicated with a superscript T. Strains with a genome sequencing project registered in GOLD [33] are in bold print and the GOLD ID is mentioned after the accession number. Published genomes are indicated with an asterisk.

Table 2. Compatibility of SRDI943 with eleven *Trifolium* genotypes for nodulation (Nod) and N₂-Fixation (Fix)

Species Name	Cultivar	Common Name	Growth Type	Nod	Fix	Reference
<i>T. glanduliferum</i> Boiss.	Prima	Gland	Annual	+	+	
<i>T. michelianum</i> Savi.	Bolta	Balansa	Annual	+	+	
<i>T. purpureum</i> Loisel	Paratta	Purple	Annual	+	+	[11]
<i>T. resupinatum</i> L.	Kyambro	Persian	Annual	+	+	
<i>T. subterraneum</i> L.	Campeda	Sub. clover	Annual	+	+	[9,11]
<i>T. subterraneum</i> L.	Clare	Sub. clover	Annual	+	+	[9,11]
<i>T. vesiculosum</i> Savi.	Arrotas	Arrowleaf	Annual	+	+	
<i>T. fragiferum</i> L.	Palestine	Strawberry	Perennial	+	+	
<i>T. polymorphum</i> Poir	Acc.#087102	Polymorphous	Perennial	+(w)	-	[11]
<i>T. pratense</i> L.	-	Red	Perennial	+(w)	-	
<i>T. repens</i> L.	Haifa	White	Perennial	+	+	

(w) indicates nodules present were white.

Table 3. Genome sequencing project information for *Rhizobium leguminosarum* bv. *trifolii* strain SRDI943.

MIGS ID	Property	Term
MIGS-31	Finishing quality	Improved high-quality draft
MIGS-28	Libraries used	2× Illumina libraries; Std short PE & CLIP long PE
MIGS-29	Sequencing platforms	Illumina HiSeq 2000
MIGS-31.2	Sequencing coverage	Illumina (761×)
MIGS-30	Assemblers	Velvet 1.1.05, phrap SPS-4.24, Allpaths version 39750
MIGS-32	Gene calling methods	Prodigal 1.4, GenePRIMP
	GOLD ID	Gi08842
	NCBI project ID	89687
	Database: IMG	2517093000
	Project relevance	Symbiotic N ₂ fixation, agriculture

Genome sequencing and assembly

The genome of *R. l. trifolii* strain SRDI943 was sequenced at the Joint Genome Institute (JGI) using an Illumina sequencing platform. An Illumina short-insert paired-end (PE) library with an average insert size of 270 bp produced 18,764,470 reads and an Illumina CLIP long-insert paired-end (PE) library with an average insert size of 9,482 bp produced 18,761,080 reads totaling 5,629 Mb of Illumina data for this genome. All general aspects of library construction and sequencing performed at the JGI can be found at the DOE JGI user homepage [35]. The initial draft assembly contained 5 contigs in 5 scaffolds. The initial draft data was assembled with Allpaths, version 39750. The Allpaths consensus was computationally shredded into 10 Kb overlapping fake reads (shreds). Illumina sequencing data were assembled with Velvet, version 1.1.05 [36], and the consensus sequences were computationally shredded into 1.5 kb overlapping fake reads (shreds). The Allpaths consensus shreds, the Illumina VELVET consensus shreds and a sub-set of the Illumina CLIP paired-end reads were integrated using parallel phrap, version SPS - 4.24 (High Performance Software, LLC). The software Consed [37-39] was used in the following finishing process. The estimated genome size is 7.4 Mb and the final assembly is based on 5,629 Mb of Illumina draft data which provides an average of 761 \times coverage of the genome.

Genome annotation

Genes were identified using Prodigal [40] as part of the DOE-JGI annotation pipeline [41] annotation pipeline, followed by a round of manual curation using the JGI GenePRIMP pipeline [42]. The predicted CDSs were translated and used to search the National Center for Biotechnology Information (NCBI) non-redundant database, UniProt, TIGRFam, Pfam, PRIAM, KEGG, COG, and InterPro databases. These data sources were combined to ascribe a product description for each predicted protein. Non-coding genes and miscellaneous features were predicted using tRNAscan-SE [43], RNAMMer [44], Rfam [45], TMHMM [46], and SignalP [47]. Additional gene prediction analyses and functional annotation were performed within the Integrated Microbial Genomes (IMG-ER) platform [35,48].

Genome properties

The genome is 7,412,387 nucleotides with 60.69% GC content (Table 4) and comprised of 5 scaffolds (Figure 3) of 5 contigs. From a total of 7,406 genes, 7,317 were protein encoding and 89 RNA only encoding genes. The majority of genes (78.5%) were assigned a putative function whilst the remaining genes were annotated as hypothetical. The distribution of genes into COGs functional categories is presented in Table 5.

Table 4. Genome Statistics for *Rhizobium leguminosarum* bv. *trifolii* SRDI943

Attribute	Value	% of Total
Genome size (bp)	7,412,387	100.00
DNA coding region (bp)	6,395,342	86.28
DNA G+C content (bp)	4,498,817	60.69
Number of scaffolds	5	
Number of contigs	5	
Total gene	7,406	100.00
RNA genes	89	1.20
rRNA operons	3	
Protein-coding genes	7,317	98.80
Genes with function prediction	5,814	78.50
Genes assigned to COGs	5,770	77.91
Genes assigned Pfam domains	6,032	81.45
Genes with signal peptides	631	8.52
Genes with transmembrane proteins	1,618	21.85
CRISPR repeats	0	

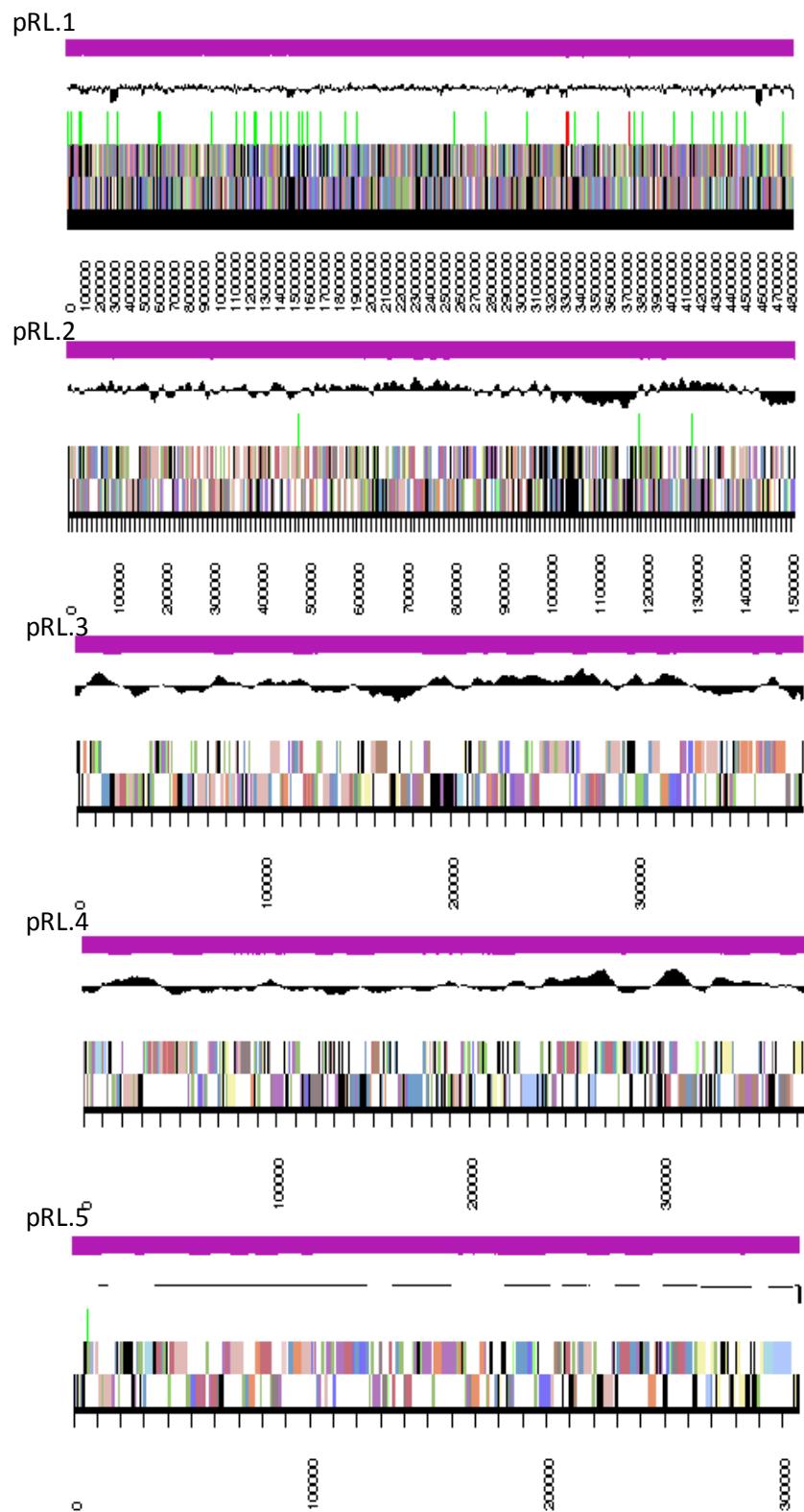


Figure 3. Graphical map of the genome of *Rhizobium leguminosarum* bv. *trifolii* strain SRDI943. From bottom to the top of each scaffold: Genes on forward strand (color by COG categories as denoted by the IMG platform), Genes on reverse strand (color by COG categories), RNA genes (tRNAs green, sRNAs red, other RNAs black), GC content, GC skew.

Table 5. Number of protein coding genes of *Rhizobium leguminosarum* bv. *trifolii* SRDI943 associated with the general COG functional categories.

Code	Value	%age	COG Category
J	196	3.03	Translation, ribosomal structure and biogenesis
A	1	0.02	RNA processing and modification
K	652	10.06	Transcription
L	231	3.57	Replication, recombination and repair
B	2	0.03	Chromatin structure and dynamics
D	40	0.62	Cell cycle control, mitosis and meiosis
Y	0	0.00	Nuclear structure
V	76	1.17	Defense mechanisms
T	373	5.76	Signal transduction mechanisms
M	334	5.16	Cell wall/membrane biogenesis
N	92	1.42	Cell motility
Z	1	0.02	Cytoskeleton
W	1	0.02	Extracellular structures
U	95	1.47	Intracellular trafficking and secretion
O	193	2.98	Posttranslational modification, protein turnover, chaperones
C	324	5.00	Energy production conversion
G	714	11.02	Carbohydrate transport and metabolism
E	659	10.17	Amino acid transport metabolism
F	109	1.68	Nucleotide transport and metabolism
H	192	2.96	Coenzyme transport and metabolism
I	227	3.50	Lipid transport and metabolism
P	333	5.14	Inorganic ion transport and metabolism
Q	165	2.55	Secondary metabolite biosynthesis, transport and catabolism
R	842	13.00	General function prediction only
S	627	9.68	Function unknown
-	1,636	22.09	Not in COGS

Acknowledgements

This work was performed under the auspices of the US Department of Energy's Office of Science, Biological and Environmental Research Program, and by the University of California, Lawrence Berkeley National Laboratory under contract No. DE-AC02-05CH11231, Lawrence Livermore National Laboratory under Contract No. DE-AC52-07NA27344, and Los Alamos National Laboratory under contract No. DE-AC02-06NA25396. We gratefully acknowledge the funding received from the Mur-

doch University Strategic Research Fund through the Crop and Plant Research Institute (CaPRI), the Centre for Rhizobium Studies (CRS) at Murdoch University and the GRDC National Rhizobium Program (UMU00032). The authors would like to thank the Australia-China Joint Research Centre for Wheat Improvement (ACCWI) and SuperSeed Technologies (SST) for financially supporting Mohamed Ninawi's PhD project.

References

1. O'Hara GW. The role of nitrogen fixation in crop production. *J Crop Prod* 1998; **1**:115-138. http://dx.doi.org/10.1300/J144v01n02_06
2. Howieson JG, Yates RJ, Foster K, Real D, Besier B. Prospects for the future use of legumes. In: Dilworth MJ, James EK, Sprent JI, Newton WE, editors. *Leguminous Nitrogen-Fixing Symbioses*. London, UK: Elsevier; 2008. p 363-394.
3. Herridge DF, Peoples MB, Boddey RM. Global inputs of biological nitrogen fixation in agricultural systems. *Plant Soil* 2008; **311**:1-18. <http://dx.doi.org/10.1007/s11104-008-9668-3>
4. Unkovich MJ, Baldock J, Peoples MB. Prospects and problems of simple linear models for estimating symbiotic N₂ fixation by crop and pasture legumes. *Plant Soil* 2010; **329**:75-89. <http://dx.doi.org/10.1007/s11104-009-0136-5>
5. Denton MD, Coventry DR, Bellotti WD, Howieson JG. Distribution, abundance and symbiotic effectiveness of *Rhizobium leguminosarum* bv. *trifolii* from alkaline pasture soils in South Australia. *Anim Prod Sci* 2000; **40**:25-35. <http://dx.doi.org/10.1071/EA99035>
6. Drew EA, Charman N, Dingemanse R, Hall E, Ballard RA. Symbiotic performance of Mediterranean *Trifolium* spp. with naturalised soil rhizobia. *Crop Pasture Sci* 2011; **62**:903-913. <http://dx.doi.org/10.1071/CP11047>
7. Rys GJ, Bonish PM. Effectiveness of *Rhizobium trifolii* populations associated with *Trifolium* species in Taranaki, New Zealand. *New Zealand Journal of Experimental Agriculture* 1981; **9**:329-335. <http://dx.doi.org/10.1080/03015521.1981.10425430>
8. Slattery JF, Coventry DR. Acid-tolerance and symbiotic effectiveness of *Rhizobium leguminosarum* bv. *trifolii* isolated from subterranean clover growing in permanent pastures. *Soil Biol Biochem* 1995; **27**:111-115. [http://dx.doi.org/10.1016/0038-0717\(94\)00143-O](http://dx.doi.org/10.1016/0038-0717(94)00143-O)
9. Drew EA, Ballard RA. Improving N₂ fixation from the plant down: Compatibility of *Trifolium subterraneum* L. cultivars with soil rhizobia can influence symbiotic performance. *Plant Soil* 2010; **327**:261-277. <http://dx.doi.org/10.1007/s11104-009-0052-8>
10. Sitepu SP. Selection of Balansa clover lines using naturalised soil rhizobia: University of Adelaide; 2001. 96 p.
11. Melino VJ, Drew EA, Ballard RA, Reeve WG, Thomson G, White RG, O'Hara GW. Identifying abnormalities in symbiotic development between *Trifolium* spp. and *Rhizobium leguminosarum* bv. *trifolii* leading to sub-optimal and ineffective nodule phenotypes. *Ann Bot (Lond)* 2012; **110**:1559-1572. [PubMed](http://dx.doi.org/10.1093/aob/mcs206) <http://dx.doi.org/10.1093/aob/mcs206>
12. Howieson JG, Ewing MA, D'antuono MF. Selection for acid tolerance in *Rhizobium meliloti*. *Plant Soil* 1988; **105**:179-188. <http://dx.doi.org/10.1007/BF02376781>
13. Field D, Garrity G, Gray T, Morrison N, Selengut J, Sterk P, Tatusova T, Thomson N, Allen M, Angiuoli SV, et al. Towards a richer description of our complete collection of genomes and metagenomes "Minimum Information about a Genome Sequence" (MIGS) specification. *Nat Biotechnol* 2008; **26**:541-547. [PubMed](http://dx.doi.org/10.1038/nbt1360) <http://dx.doi.org/10.1038/nbt1360>
14. Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc Natl Acad Sci USA* 1990; **87**:4576-4579. [PubMed](http://dx.doi.org/10.1073/pnas.87.12.4576) <http://dx.doi.org/10.1073/pnas.87.12.4576>
15. Garrity GM, Bell JA, Lilburn T. Phylum XIV. *Proteobacteria* phyl. nov. In: Garrity GM, Brenner DJ, Krieg NR, Staley JT (eds), *Bergey's Manual of*

Systematic Bacteriology, Second Edition, Volume 2, Part B, Springer, New York, 2005, p. 1.

16. Garrity GM, Bell JA, Lilburn T. Class I. *Alphaproteobacteria* class. nov. In: Garrity GM, Brenner DJ, Krieg NR, Staley JT (eds), Bergey's Manual of Systematic Bacteriology, Second Edition, Volume 2, Part C, Springer, New York, 2005, p. 1.

17. Validation List No. 107. List of new names and new combinations previously effectively, but not validly, published. *Int J Syst Evol Microbiol* 2006; **56**:1-6. [PubMed](#) <http://dx.doi.org/10.1099/ij.s.0.64188-0>

18. Kuykendall LD. Order VI. *Rhizobiales* ord. nov. In: Garrity GM, Brenner DJ, Krieg NR, Staley JT, editors. Bergey's Manual of Systematic Bacteriology. Second ed: New York: Springer - Verlag; 2005. p 324.

19. Kuykendall LD. Family I. *Rhizobiaceae* In: Garrity GM, Brenner DJ, Krieg NR, Staley JT, editors. Bergey's Manual of Systematic Bacteriology. New York: Springer - Verlag; 2005.

20. Conn HJ. Taxonomic relationships of certain non-sporeforming rods in soil. *J Bacteriol* 1938; **36**:320-321.

21. Skerman VBD, McGowan V, Sneath PHA. Approved Lists of Bacterial Names. *Int J Syst Bacteriol* 1980; **30**:225-420. [http://dx.doi.org/10.1099/00207713-30-1-225](#)

22. Kuykendall LD, Young JM, Martínez-Romero E, Kerr A, Sawada H. Genus I. *Rhizobium*. In: Garrity GM, Brenner DJ, Krieg NR, Staley JT, editors. Bergey's Manual of Systematic Bacteriology. Second ed. Volume 2. New York: Springer - Verlag; 2005.

23. Frank B. Über die Pilzsymbiose der Leguminosen. *Ber Dtsch Bot Ges* 1889; **7**:332-346.

24. Jordan DC, Allen ON. Genus I. *Rhizobium* Frank 1889, 338; Nom. gen. cons. Opin. 34, Jud. Comm. 1970, 11. In: Buchanan RE, Gibbons NE (eds), Bergey's Manual of Determinative Bacteriology, Eighth Edition, The Williams and Wilkins Co., Baltimore, 1974, p. 262-264.

25. Young JM, Kuykendall LD, Martínez-Romero E, Kerr A, Sawada H. A revision of *Rhizobium* Frank 1889, with an emended description of the genus, and the inclusion of all species of *Agrobacterium* Conn 1942 and *Allorhizobium undicola* de Lajudie et al. 1998 as new combinations: *Rhizobium radiobacter*, *R. rhizogenes*, *R. rubi*, *R. undicola* and *R. vitis*. *Int J Syst Evol Microbiol* 2001; **51**:89-103. [PubMed](#)

26. Editorial Secretary (for the Judicial Commission of the International Committee on Nomenclature of Bacteria)

27. Ramírez-Bahena MH, García-Fraile P, Peix A, Valverde A, Rivas R, Igual JM, Mateos PF, Martínez-Molina E, Velázquez E. Revision of the taxonomic status of the species *Rhizobium leguminosarum* (Frank 1879) Frank 1889AL, *Rhizobium phaseoli* Dangeard 1926AL and *Rhizobium trifolii* Dangeard 1926AL. *R. trifolii* is a later synonym of *R. leguminosarum*. Reclassification of the strain *R. leguminosarum* DSM 30132 (=NCIMB 11478) as *Rhizobium pisi* sp. nov. *Int J Syst Evol Microbiol* 2008; **58**:2484-2490. [PubMed](#) <http://dx.doi.org/10.1099/ij.s.0.65621-0>

28. Editorial Secretary (for the Judicial Commission of the International Committee on Nomenclature of Bacteria). OPINION 34: Conservation of the Generic Name *Rhizobium* Frank 1889. *Int J Syst Bacteriol* 1970; **20**:11-12. [http://dx.doi.org/10.1099/00207713-20-1-11](#)

29. Agents B. Technical rules for biological agents. TRBA (<http://www.baua.de>):466.

30. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 2000; **25**:25-29. [PubMed](#) <http://dx.doi.org/10.1038/75556>

31. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Mol Biol Evol* 2011; **28**:2731-2739. [PubMed](#) <http://dx.doi.org/10.1093/molbev/msr121>

32. Felsenstein J. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 1985; **39**:783-791. [http://dx.doi.org/10.2307/2408678](#)

33. Liolios K, Mavromatis K, Tavernarakis N, Kyrpides NC. The Genomes On Line Database (GOLD) in 2007: status of genomic and metagenomic projects and their associated metadata. *Nucleic Acids Res* 2008; **36**:D475-D479. [PubMed](#) <http://dx.doi.org/10.1093/nar/gkm884>

34. Reeve WG, Tiwari RP, Worsley PS, Dilworth MJ, Glenn AR, Howieson JG. Constructs for insertional mutagenesis, transcriptional signal localization and gene regulation studies in root nodule and other bacteria. *Microbiology* 1999;

145:1307-1316. [PubMed](#)
<http://dx.doi.org/10.1099/13500872-145-6-1307>

35. DOE Joint Genome Institute.
<http://my.jgi.doe.gov/general/index.html>

36. Zerbino DR. Using the Velvet *de novo* assembler for short-read sequencing technologies. Current Protocols in Bioinformatics 2010;Chapter 11:Unit 11.5.

37. Ewing B, Green P. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res* 1998; **8**:186-194. [PubMed](#)
<http://dx.doi.org/10.1101/gr.8.3.175>

38. Ewing B, Hillier L, Wendl MC, Green P. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res* 1998; **8**:175-185. [PubMed](#)
<http://dx.doi.org/10.1101/gr.8.3.175>

39. Gordon D, Abajian C, Green P. Consed: a graphical tool for sequence finishing. *Genome Res* 1998; **8**:195-202. [PubMed](#)
<http://dx.doi.org/10.1101/gr.8.3.195>

40. Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 2010; **11**:119. [PubMed](#) <http://dx.doi.org/10.1186/1471-2105-11-119>

41. Mavromatis K, Ivanova NN, Chen IM, Szeto E, Markowitz VM, Kyrpides NC. The DOE-JGI Standard operating procedure for the annotations of microbial genomes. *Stand Genomic Sci* 2009; **1**:63-67. [PubMed](#)
<http://dx.doi.org/10.4056/sigs.632>

42. Pati A, Ivanova NN, Mikhailova N, Ovchinnikova G, Hooper SD, Lykidis A, Kyrpides NC.

GenePRIMP: a gene prediction improvement pipeline for prokaryotic genomes. *Nat Methods* 2010; **7**:455-457. [PubMed](#)
<http://dx.doi.org/10.1038/nmeth.1457>

43. Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* 1997; **25**:955-964. [PubMed](#)

44. Lagesen K, Hallin P, Rodland EA, Staerfeldt HH, Rognes T, Ussery DW. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res* 2007; **35**:3100-3108. [PubMed](#)
<http://dx.doi.org/10.1093/nar/gkm160>

45. Griffiths-Jones S, Bateman A, Marshall M, Khanna A, Eddy SR. Rfam: an RNA family database. *Nucleic Acids Res* 2003; **31**:439-441. [PubMed](#)
<http://dx.doi.org/10.1093/nar/gkg006>

46. Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* 2001; **305**:567-580. [PubMed](#)
<http://dx.doi.org/10.1006/jmbi.2000.4315>

47. Bendtsen JD, Nielsen H, von Heijne G, Brunak S. Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol* 2004; **340**:783-795. [PubMed](#)
<http://dx.doi.org/10.1016/j.jmb.2004.05.028>

48. Markowitz VM, Mavromatis K, Ivanova NN, Chen IM, Chu K, Kyrpides NC. IMG ER: a system for microbial genome annotation expert review and curation. *Bioinformatics* 2009; **25**:2271-2278. [PubMed](#)
<http://dx.doi.org/10.1093/bioinformatics/btp393>